Surface enhanced Raman spectroscopy using a single mode nanophotonic-plasmonic platform

Frédéric Peyskens,[*,†] Ashim Dhakal,[†] Pol Van Dorpe,[‡] Nicolas Le Thomas,[†] and Roel Baets[†]

Photonics Research Group, INTEC-department, Ghent University-imec; Center for Nano-and BioPhotonics, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium, and imec, Kapeldreef 75, 3001 Heverlee, Belgium; Department of Physics, KULeuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

E-mail: fpeysken@intec.ugent.be

---

[*]To whom correspondence should be addressed

[†]Photonics Research Group, INTEC-department, Ghent University-imec; Center for Nano-and BioPhotonics, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

[‡]imec, Kapeldreef 75, 3001 Heverlee, Belgium; Department of Physics, KULeuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

Surface Enhanced Raman Spectroscopy (SERS) is a well-established technique for enhancing Raman signals.[1–12] Recently photonic integrated circuits have been used, as an alternative to microscopy based excitation and collection, to probe SERS signals from external metallic nanoparticles.[13–15] However, in order to develop quantitative on-chip SERS sensors, integration of dedicated nanoplasmonic antennas and waveguides[16–22] is desirable. Here we bridge this gap by demonstrating for the first time the generation of SERS signals from integrated bowtie nanoantennas, excited and collected by a single mode waveguide, and rigorously quantify the enhancement process. The guided Raman power generated by a 4-Nitrothiophenol coated bowtie antenna shows an 8×10^6 enhancement compared to the free-space Raman scattering. An excellent correspondence is obtained between the theoretically predicted and observed absolute Raman power. This work paves the way towards fully integrated lab-on-a-chip systems where the single mode SERS-probe can be combined with other photonic, fluidic or biological functionalities.

A schematic of the device under study is shown in Figure 1(a). The fundamental TE-mode of a silicon nitride (SiN) rib waveguide excites a periodic array of gold bowtie antennas coated with a 4-Nitrothiophenol (NTP) monolayer. The pump wavelength for all experiments is set to $\lambda_P = 785$ nm and NTP Stokes light (at λ_S) is subsequently collected back into the same waveguide mode. Fabrication details can be found in the Methods section and a description of the measurement setup is outlined in the Supplementary Information S1. A scanning electron microscope image of the functionalized waveguide, with cross-sectional area of 220×700 nm^2, is depicted in Figure 1(b). Raman spectra of an uncoated and coated waveguide functionalized with 40 antennas are shown in Figure 1(c). The spectral regions where an NTP Stokes peak is expected (1,080, 1,110, 1,340 and 1,575 cm^{-1})[23] are highlighted by the cyan shaded areas. Before coating no NTP peaks can be distinguished from the inherent SiN background. The peaks at 1,250 and 1,518 cm^{-1} (marked by the black dashed lines) are attributed to interference effects of the Au array which act on the scattered background light, and they are also observed on the extinction curves of the functionalized waveguides (see Supporting Information S2). Hence they do not represent specific

2

Raman lines. After coating, four additional peaks appear and coincide with the expected NTP Stokes peaks. This demonstrates that SERS signals from single monolayer coated antennas can be efficiently excited and collected by the same fundamental waveguide mode.

Subsequently the dependence of the SERS signal on the position of the plasmon resonance was investigated to verify that it can be attributed to a resonance effect and not to coincidental surface roughness. To this end, waveguides functionalized with a fixed number of antennas but varying bowtie geometries were considered. The relevant bowtie parameters are its length L, gap Δ and apex angle α (Figure 2(a)). By changing the length, the antenna resonance can be tuned ($L_1 = 90$ nm, $L_2 = 115$ nm and $L_3 = 140$ nm for fixed $\alpha = 60°$ and $\Delta = 40$ nm). Extinction spectra are plotted in Figure 2(b) while the corresponding Raman spectra are depicted in Figure 2(c). The Raman spectrum of a reference waveguide without any Au functionalization is also shown. Even after coating the reference waveguide does not generate NTP peaks, so any Raman signal indeed originates from the antenna region and does not contain contributions from spontaneous Raman scattering along the waveguide.[24] The L_1 resonance is detuned from the relevant pump and Stokes region, resulting in a poor Raman spectrum. By increasing the length (L_2 and L_3 bowties) the resonance redshifts and lines up with the pump and Stokes wavelengths. For these bowties the NTP spectrum starts to emerge. The reported SERS spectra can hence be attributed to a plasmon resonance effect such that a stable and reproducible enhancement factor can be associated with them, in contrast to SERS events originating from random surface defects. The increased overlap with the plasmon resonance, and hence extinction, also results in a decreased background.

Due to the metal induced loss, there will exist an optimum number N_{opt} of patterned antennas such that the SERS signal is maximized. Such an optimum is investigated in Figure 3 for a fixed bowtie geometry ($\alpha = 60°$, $L = 100$ nm and $\Delta = 40$ nm) but varying N: $N = 10, 20, 30, 40, 70$ and $N = 0$ which is a reference waveguide. Each waveguide is measured 10 times and the averaged Raman spectra are reported in Figure 3(a). The $N = 70$ signal is not shown since it could not be distinguished from the inherent offset signal of the detector. For a given fixed input power, corresponding to roughly 5 mW guided power, the reference waveguide generates a considerable

background signal in the 1,340 cm^{-1} region, where the strongest NTP peak is expected. Functionalizing the waveguide with increasing N reduces this unwanted background due to the attenuation caused by the nanoantennas. In addition the 1,340 cm^{-1} peak starts to emerge when N increases. The smaller peaks at 1,080, 1,110 and 1,575 cm^{-1} only appear when the background is sufficiently low. A zoom on the dominant 1,340 cm^{-1} peak (cyan dashed line) is shown in Figure 3(b). For clarity the background is locally subtracted. As expected, the signal reaches a maximum value for $10 \leq N \leq 20$ and then decays again with increasing N. Apart from signal optimization it is however equally important to reduce the SiN background in order to resolve the smallest spectral features. Therefore an analytical model is developed to outline the interplay between signal enhancement and background reduction, and to derive the relevant figure of merit for on-chip SERS.

Figure 4(a) shows a schematic longitudinal cross-section of the chip consisting of N antennas spaced with period $\Lambda = 10 \ \mu$m. Each array is centered on the waveguide with a distance $L_1 \approx 0.5$ cm to the front and back facet of the chip. The NTP monolayer on each antenna will generate a forward propagating Stokes power $P_A(\lambda_P, \lambda_S)$ for a given pump power P_{pump}. This single antenna conversion efficiency $P_A(\lambda_P, \lambda_S)$ is an antenna dependent factor incorporating the integrated field enhancement profile near the metal surface and the molecular density and Raman cross-section. The total transmission loss induced by one antenna at wavelength λ is given by $1 - e_{\lambda}^{-1}$, whereby e_{λ} is the linear antenna extinction. Apart from the intrinsic waveguide losses α_{wg}, the pump and Stokes light will hence also be attenuated by e_P and e_S respectively. In the Supplementary Information S3 it is then shown that the total Stokes power P_{tot} generated by an array of N coated antennas is approximately given by

$$\frac{P_{tot}}{P_{pump}} \approx P_A(\lambda_P, \lambda_S) \mathrm{e}^{-2\alpha_{wg}L_1} e_S^{(1-N)} \left(\frac{1 - \left(\frac{e_S}{e_P}\right)^N}{1 - \left(\frac{e_S}{e_P}\right)} \right) = FOM(N, \lambda_P, \lambda_S) \mathrm{e}^{-2\alpha_{wg}L_1}.$$

The quantity $FOM(N, \lambda_P, \lambda_S)$ contains all necessary parameters to assess the SERS signal strength for a given waveguide geometry and is hence considered to be the relevant figure of merit (FOM) in comparing the performance of integrated antenna arrays. The optimum antenna number $N_{opt} =$

$\log\left\{\log\left(e_S\right)/\log\left(e_P\right)\right\}/\log\left(e_S/e_P\right)$. For the particular bowtie antenna studied in Figure 3, the extinction spectrum $e(\lambda)$ and single antenna conversion efficiency $P_A(\lambda_P, \lambda_S)$ are numerically evaluated using Lumerical FDTD Solutions (see Methods section). The predicted N_{opt} for the 1,340 cm^{-1} peak is 10 antennas, using the simulated extinctions $e_P = 1.14$ ($E_P = 0.58$ dB) and $e_S = 1.08$ ($E_S = 0.34$ dB), while $P_A \approx 2.35 \times 10^{-15}$. For each 1W of pump power the antenna is therefore expected to generate 2.35 fW of guided Stokes power. The Raman enhancement factor is calculated through $EF_R = \beta(\lambda_P)^2 \beta(\lambda_S)^2$ in which $\beta(\lambda)$ is the local field enhancement (see Methods section). In the center of the gap at 5 nm from the tip of the antenna (marked by the black dot in Figure 2(a)) an $EF_R \approx 1.42 \times 10^4$ is expected . Apart from the relevant NTP signal, the SiN itself generates a considerable background while the pump beam is propagating along the waveguide. This background signal P_{bg} can be approximated by

$$\frac{P_{bg}}{P_{pump}} \approx P_B(\lambda_P, \lambda_S) e^{-2\alpha_{wg}L_1} \left(e_P^{-N} + e_S^{-N}\right)$$

in which $P_B(\lambda_P, \lambda_S)$ is a waveguide dependent factor incorporating the specific modal field profile and the SiN molecular density and cross-section (see Supplementary Information S4).

Our analytical model and the associated numerical calculations will now be benchmarked against the spectra from Figure 3 to verify whether the theoretically estimated power values correspond to the experimentally obtained absolute Raman powers. To this end, the NTP signal strength at 1,340 cm^{-1}, obtained from Figure 3(b), is analyzed as a function of the antenna number N and compared with the theoretical estimations. Furthermore, the background associated shot noise is also calculated. The results are depicted in Figure 4(b). While the ideal model assumes N identical antennas, the fabricated antennas will show differences among each other resulting in changes of e_P, e_S and P_A from one antenna to the other. A generalized model incorporating potential differences in e_P, e_S and P_A is described in the Supplementary Information S5. In order to estimate the uncertainty on these experimental parameters a randomized fit to the generalized model has been applied. A set of normally distributed numbers is generated for each of the three

parameters and then plugged into the generalized model to calculate the distribution of signal and shot noise counts, defining an area within which the probable signal (blue area) and noise (red area) counts are situated. The mean values of these distributions are extracted from an initial constrained fit to the ideal model (dotted lines), while its standard deviations are chosen such that the corresponding signal and shot noise distributions cover all experimental datapoints (red and blue dots). Based on the randomized fit it is possible to estimate the spread on the experimental parameters: $E_P \approx 0.49 \pm 0.11$ dB, $E_S \approx 0.35 \pm 0.11$ dB and $P_A = (2.60 \pm 0.77) \times 10^{-15}$ (theoretically $E_P = 0.58$ dB, $E_S = 0.34$ dB and $P_A = 2.35 \times 10^{-15}$ were predicted). The theoretically predicted parameters are all within the error bars of the experimentally fitted data, which clearly establishes the validity of our model and its ability to provide quantitative predictions of the absolute Raman power coupled into a single mode waveguide. Given this excellent correspondence, we expect the fabricated structures to have a Raman enhancement factor EF_R on the order of 1.42×10^4 near the two antenna gap tips. Decreasing the gap size should boost EF_R and P_A by another two or three orders of magnitude. From the fitting values the optimum antenna number is estimated to be 11 ± 3 (compared to 10 theoretically). The single antenna conversion efficiency P_A shows that the fabricated antennas produce (2.60 ± 0.77) fW of guided Stokes power for each 1W of guided pump power. Compared to the free space Raman scattering P_0 of a single NTP molecule in a bulk air environment $P_A \approx 3.94 \times 10^6 P_0$. This includes the excitation and emission enhancement of all molecules in the monolayer as well as the coupling efficiency to the guided mode. Since only half of the Stokes light is carried by the forward propagating mode, the total power coupled into the fundamental TE-mode is therefore $\approx 8 \times 10^6 P_0$.

Our observations also reveal that a minimum number of antennas N_{min} is required to generate a detectable signal (marked by the white square in Figure 4(b)). If $N < N_{min}$ then the shot noise still dominates on the signal. It has to be noted however that the relevant signal is generated in a very small region $(N-1)\Lambda$ compared to the overall length $2L_1 + (N-1)\Lambda \approx 2L_1$, while the shot noise is mainly attributed to this non-useful length $2L_1$. Chip designs which allow a separation of the signal from the background are expected to have $N_{min} = 1$ such that signals originating from one

single antenna can still be detected. As a result it would become possible to simultaneously probe large areas of analytes ($> \lambda^2$) and detecting all SERS events, originating from different locations, by monitoring just a single waveguide output in contrast to microscopy based systems where one has to serially scan all hotspot locations.

The work presented here paves the way towards the efficient design of evanescently coupled nanoantennas for on-chip excitation and emission enhancement in the 700-1000 nm region. Due to the low fluorescence, negligible water absorption and the availability of high quality and low-cost sources and detectors this region is of particular interest for Raman sensing.[28] In combination with other on-chip spectral functionalities, such as arrayed waveguide gratings,[28] the presented platform is forecasted to allow multiplexed detection of extremely weak Raman signals on a highly dense integrated platform. We also envisage that integrated nanoantennas, similar to the ones reported here, could be used as transducer between quantum dot emitters and the fundamental waveguide mode, potentially enabling applications in on-chip quantum communication and quantum computation.[29,30]

Methods

Fabrication details

The fabrication consists of a 2-step e-beam lithography process. In the first step the nanoplasmonic antennas are patterned on top of a slab Si/SiO$_2$/SiN wafer using a positive PMMA e-beam resist. After PMMA exposure, the samples are developed in a 1:1 MIBK:IPA solution after which a 2 nm Ti adhesion layer and 30 nm Au layer are deposited in a commercial Pfeiffer Spider sputter system. The samples are then immersed in acetone for lift-off. In the second step the waveguides are defined. After metal lift-off a negative ma-N 2403 resist is spun, exposed and developed in ma-D 525. An e-spacer is also spun on top of the ma-N 2403 to avoid charging effects. The developed samples are then etched with an ICP plasma (C$_4$F$_8$/SF$_6$ mixture) in a commercial Oxford Plasmalab system. After resist strip and cleaning, the samples are immersed overnight in a 1 mM NTP:EtOH

solution and subsequently rinsed with pure ethanol to remove the residual NTP. A self-assembled monolayer of NTP is assumed to form on the Au surface through a Au-S bond.[23]

Numerical Simulations

Numerical simulations were performed with Lumerical FDTD Solutions. We used a refractive index of $n_{rib} = 1.9$ for the SiN rib (with width $w_{rib} = 700$ nm and height $h_{rib} = 220$ nm), $n_{uclad} = 1.45$ for the SiO_2 undercladding and $n_{tclad} = 1$ for the top cladding (air). The Si substrate was not taken into account since the real oxide cladding is thick enough to avoid substantial power leakage to the Si such that the numerical results faithfully represent the actual experimental conditions. A thin native oxide layer ($t_{nox} = 2$ nm) between the SiN and the Ti has also been incorporated.[19] The metal stack thicknesses are fixed to $t_{Ti} = 2$ nm and $t_{Au} = 30$ nm and a built-in refractive index model for Au (Johnson and Christy[25]) and Ti (CRC[26]) is used. An additional surface layer with thickness $t_{NTP} = 1$ nm and index $n_{NTP} = 3$ is used to model the NTP monolayer. The antenna region (including the Ti adhesion layer and the NTP monolayer) is meshed with a uniform mesh of 0.5 nm in the plane of the antenna (yz-plane) and 2 nm in the x-direction. A mesh refinement to 1 nm is applied in regions where the thicknesses in the x-direction are ≤ 2 nm. The estimated surface area of an NTP molecule is 0.18 nm^2, so the surface density is then $\rho_s = 5.56 \times 10^{18}$ molecules/m^2.[23] The Raman cross section is $\sigma \approx 0.358 \times 10^{-30}$ cm^2/sr, which was obtained by applying the λ_S^{-4} scaling to the original data of the 1,340 cm^{-1} line.[27] Single antenna extinction spectra $E(\lambda)$ (in dB) are calculated through $E(\lambda) = T_{ref}(\lambda) - T_{ant}(\lambda)$ in which $T_{ref}(\lambda)$ is the power transmission (in dB) through the reference waveguide and $T_{ant}(\lambda)$ the power transmission (in dB) of a waveguide functionalized with one antenna. Linear extinction spectra $e(\lambda) \stackrel{\Delta}{=} e_\lambda$ are then given by $e(\lambda) = 10^{E(\lambda)/10}$. A field and index profile monitor are used to extract the local field $|\mathbf{E}(\mathbf{r},\lambda)|$ and index $n(\mathbf{r})$ around the antenna. The single antenna conversion efficiency

$$P_A(\lambda_P, \lambda_S) = \frac{\rho_s \sigma}{2t_{NTP}} \frac{\int \int \int_{V_m} n_g(\lambda_P) n_g(\lambda_S) \lambda_S^2 |\mathbf{E}(\mathbf{r},\lambda_P)|^2 |\mathbf{E}(\mathbf{r},\lambda_S)|^2 d\mathbf{r}}{\left(\int \int n(\mathbf{r})^2 |\mathbf{E}^m(\mathbf{r},\lambda_P)|^2 d\mathbf{r} \right) \left(\int \int n(\mathbf{r})^2 |\mathbf{E}^m(\mathbf{r},\lambda_S)|^2 d\mathbf{r} \right)}$$

8

is calculated by integrating the local fields over the effective monolayer volume V_m in which the index satisfies $n(\mathbf{r})\,|_{\mathbf{r}\in V_m} = n_{NTP}$ (see Supplementary Information S3). The group index of the waveguide mode is $n_g(\lambda)$. The denominator is calculated using the modal fields $\mathbf{E}^m(\mathbf{r},\lambda)$ of a non-functionalized reference waveguide and the local field enhancement is given by the ratio of the local and modal electric fields: $\beta(\mathbf{r},\lambda) = \frac{|\mathbf{E}(\mathbf{r},\lambda)|}{|\mathbf{E}^m(\mathbf{r},\lambda)|}$. At a certain position, the Raman enhancement factor EF_R is calculated as $EF_R = \beta(\lambda_P)^2\beta(\lambda_S)^2$. Numerically calculated values are mentioned in the main text.

References

(1) Anker, J. N., et al. Biosensing with plasmonic nanosensors. *Nat. Mater.* **7,** 442–453 (2008).

(2) Willets, K. A. & Van Duyne, R.P. Localized surface plasmon resonance spectroscopy and sensing. *Annu. Rev. Phys. Chem.* **58,** 267–297 (2007).

(3) Halas, N. J., Lal, S., Chang, W-S., Link, S. & Nordlander, P. Plasmons in strongly coupled metallic nanostructures. *Chem. Rev.* **111,** 3913–3961 (2011).

(4) Giannini, V., Fernández-Domínguez, A. I., Heck, S. C. & Maier, S. A. Plasmonic nanoantennas: fundamentals and their use in controlling the radiative properties of nanoemitters. *Chem. Rev.* **111,** 3888–3912 (2011).

(5) Chu, Y., Banaee, M. G. & Crozier, K. B. Double-resonance plasmon substrates for surface-enhanced Raman scattering with enhancement at excitation and Stokes frequencies. *ACS Nano* **4 (5),** 2804–2810 (2010).

(6) McFarland, A. D., Young, M. A., Dieringer, J. A. & Van Duyne, R. P. Wavelength-scanned surface-enhanced Raman excitation spectroscopy. *J. Phys. Chem. B* **109,** 11279–11285 (2005).

(7) Ye, J., et al. Plasmonic Nanoclusters: Near Field Properties of the Fano Resonance Interrogated with SERS. *Nano Lett.* **12 (3),** 1660–1667 (2012).

(8) Kasera, S., Biedermann, F., Baumberg, J.J., Scherman, O.A. & Mahajan, S. Quantitative SERS Using the Sequestration of Small Molecules Inside Precise Plasmonic Nanoconstructs. *Nano Lett.* **12 (11),** 5924–5928 (2012).

(9) Gallinet, B., Siegfried, T., Sigg, H., Nordlander, P. & Martin, O.J.F. Plasmonic Radiance: Probing Structure at the Angström Scale with Visible Light. *Nano Lett.* **13 (2),** 497–503 (2013).

(10) Siegfried, T., Ekinci, Y., Martin, O.J.F. & Sigg, H. Gap Plasmons and Near-Field Enhancement in Closely Packed Sub- 10 nm Gap Resonators. *Nano Lett.* **13 (11),** 5449–5453 (2013).

(11) Li, J., et al. 300 mm Wafer-level, ultra-dense arrays of Au-capped nanopillars with sub-10 nm gaps as reliable SERS substrates. *Nanoscale* **6,** 12391–12396 (2014).

(12) Seok, T. J., Jamshidi, A., Eggleston, M. & Wu, M.C. Mass-producible and efficient optical antennas with CMOS-fabricated nanometer-scale gap. *Opt. Express* **21 (14),** 16561–16569 (2013).

(13) Lin, S., Zhu, W., Jin, Y. & Crozier, K.B. Surface-Enhanced Raman Scattering with Ag Nanoparticles Optically Trapped by a Photonic Crystal Cavity. *Nano Lett.* **13 (2),** 559Ű-563 (2013).

(14) Kong, L., Lee, C., Earhart, C.M., Cordovez, B. & Chan, J.W. A nanotweezer system for evanescent wave excited surface enhanced Raman spectroscopy (SERS) of single nanoparticles. *Opt. Express* **23 (5),** 6793Ű-6802 (2015).

(15) Measor, P., et al. On-chip surface-enhanced Raman scattering detection using integrated liquid-core waveguides. *Appl. Phys. Lett.* **90,** 211107-1Ű-211107-3 (2007).

(16) Peyskens, F., et al. Bright and dark plasmon resonances of nanoplasmonic antennas evanescently coupled with a silicon nitride waveguide. *Opt. Express* **23 (3),** 3088–3101 (2015).

(17) Arnaud, L., et al. Waveguide-coupled nanowire as an optical antenna. *J. Opt. Soc. Am. A* **30 (11),** 2347–2355 (2013).

(18) Arango, F. B., Kwadrin, A. & Koenderink, A. F. Plasmonic Antennas Hybridized with Dielectric Waveguides. *ACS Nano* **6 (11),** 10156–10167 (2012).

(19) Février, M., et al. Giant Coupling Effect between Metal Nanoparticle Chain and Optical Waveguide. *Nano Lett.* **12,** 1032–1037 (2012).

(20) Février, M., Gogol, P., Lourtioz, J-M. & Dagens, B. Metallic nanoparticle chains on dielectric waveguides: coupled and uncoupled situations compared. *Opt. Express* **21 (21),** 24504–24513 (2013).

(21) Février, M., et al. Integration of short gold nanoparticles chain on SOI waveguide toward compact integrated bio-sensors. *Opt. Express* **20 (16),** 17402–17410 (2012).

(22) Chamanzar, M., Xia, Z., Yegnanarayanan, S. & Adibi, A. Hybrid integrated plasmonic-photonic waveguides for on-chip localized surface plasmon resonance (LSPR) sensing and spectroscopy. *Opt. Express* **21 (26),** 32086–32098 (2013).

(23) Mahmoud, M.A. Surface-Enhanced Raman Spectroscopy of Double-Shell Hollow Nanoparticles: Electromagnetic and Chemical Enhancements. *Langmuir* **29,** 6253–6261 (2013).

(24) Dhakal, A., et al. Evanescent excitation and collection of spontaneous Raman spectra using silicon nitride nanophotonic waveguides. *Opt. Lett.* **39 (13),** 4025–4028 (2014).

(25) Johnson, P.B., Christy, R.W. Optical Constants of the Noble Metals. *Phys. Rev. B* **6 (12),** 4370–4379 (1972).

(26) Haynes, W. M., Eds. *CRC Handbook Of Chemistry And Physics* 95th Edition (Internet Version 2015) (CRC/Taylor and Francis, 2015).

(27) Thomas, M., et al. Distinguishing chemical and electromagnetic enhancement in surface-enhanced Raman spectra: The case of para-nitrothiophenol. *J. Raman Spectrosc.* **44,** 1497Ű-1505 (2013).

(28) Martens, D., et al. Compact silicon nitride arrayed waveguide gratings for very near-infrared wavelengths. *IEEE Photon. Technol. Lett.* **27(2),** 137Ű-140 (2015).

(29) Pelton, M. Modified spontaneous emission in nanophotonic structures. *Nature Photon.* **9,** 427Ű-435 (2015).

(30) Tame, M.S., et al. Quantum plasmonics. *Nature Phys.* **9,** 329-340 (2013).

(31) Chul Jun, Y., Briggs, R.M., Atwater, H.A. & Brongersma, M.L. Broadband enhancement of light emission in silicon slot waveguides. *Opt. Express* **17 (9),** 7479–7490 (2009).

(32) Long, D.A. *The Raman Effect: A Unified Treatment Of The Theory Of Raman Scattering By Molecules* pp 96-97 (John Wiley & Sons Ltd, Chichester, England, 2002).

Acknowledgement

Author Contributions

F.P. fabricated the samples, performed the numerical simulations and experiments, analyzed the data, developed the analytical model and wrote the paper. A.D. built the experimental setup and developed the measurement procedure. P.V.D., N.L.T. and R.B. supervised the work. All authors provided feedback on the manuscript.

Competing financial interests

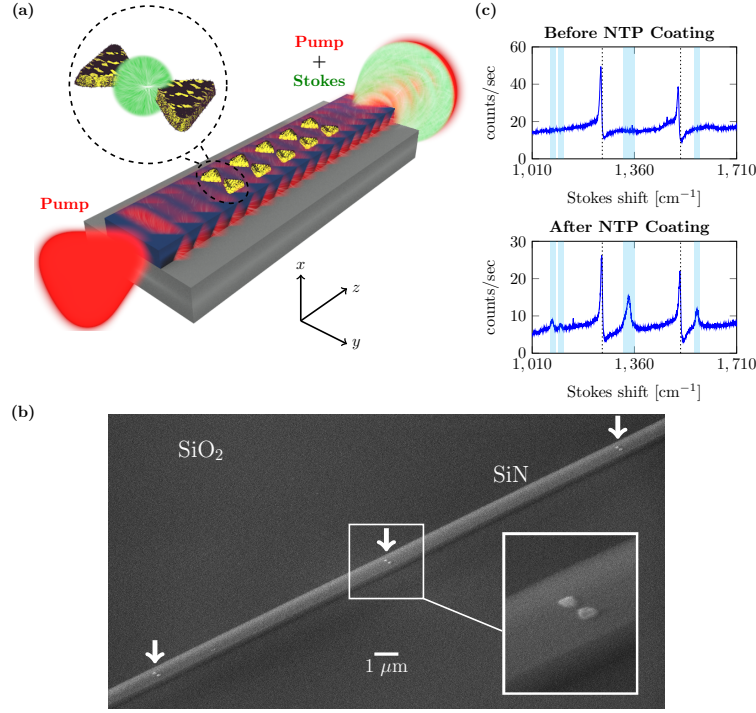The authors declare no competing financial interests.

Figure 1: **Evanescent excitation and collection of SERS spectra. (a)** Schematic of the chip consisting of single mode SiN waveguides (blue) on an SiO$_2$ undercladding (gray), functionalized with an array of gold bowtie antennas (yellow). All antennas are coated with an NTP monolayer (purple dots), evanescently excited by the fundamental TE-mode (red). The NTP Stokes signal (green) is collected by the same mode. **(b)** Scanning electron microscope image of a functionalized waveguide. The white arrows indicate antenna positions. The inset shows a zoom of a typical antenna. **(c)** Raman spectra of a waveguide functionalized with 40 antennas. The cyan shaded areas mark the NTP Stokes peaks while the black dashed lines represent peaks attributed to interference effects of the plasmonic array. Before coating, the waveguide already generates a Raman background in itself (top). After coating, NTP peaks emerge (bottom).
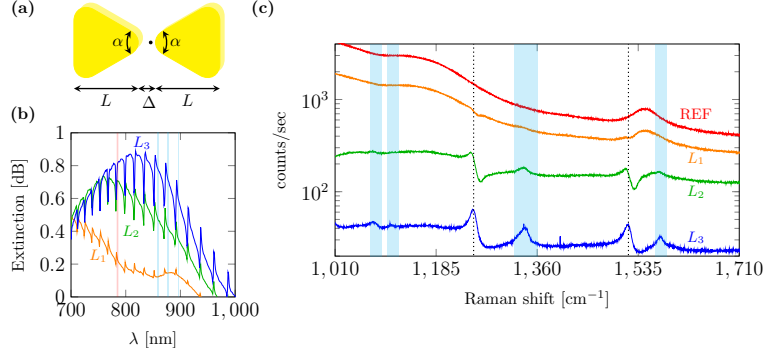
Figure 1

Figure 2: **Signal dependence on the plasmon resonance.(a)** Bowtie antenna geometrical paramters: length L, gap Δ and apex angle α. **(b)** Single antenna extinction spectra for 3 different bowtie antennas with fixed apex angle $\alpha = 60°$ and gap $\Delta = 40$ nm but varying length: $L_1 = 90$ nm (orange), $L_2 = 115$ nm (green), $L_3 = 140$ nm (blue). The red and cyan shaded lines correspond to the pump and Stokes wavelengths respectively. **(c)** Corresponding Raman spectra of the waveguides functionalized with these 3 bowtie anntenas (orange,green,blue) and Raman spectrum of the reference waveguide (red).
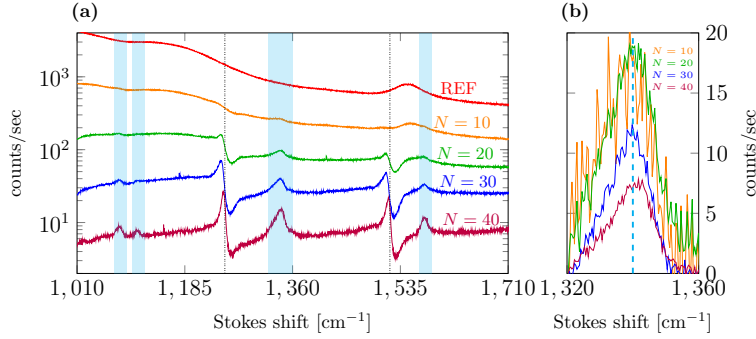
Figure 2

Figure 3: **Signal dependence on the number of antennas N. (a)** Raman spectra of a reference waveguide (REF) and waveguides functionalized with $N = 10, 20, 30, 40$ antennas. **(b)** Zoom on the 1,340 cm^{-1} signal peak. The background is locally subtracted to obtain the pure NTP signal.

Figure 3
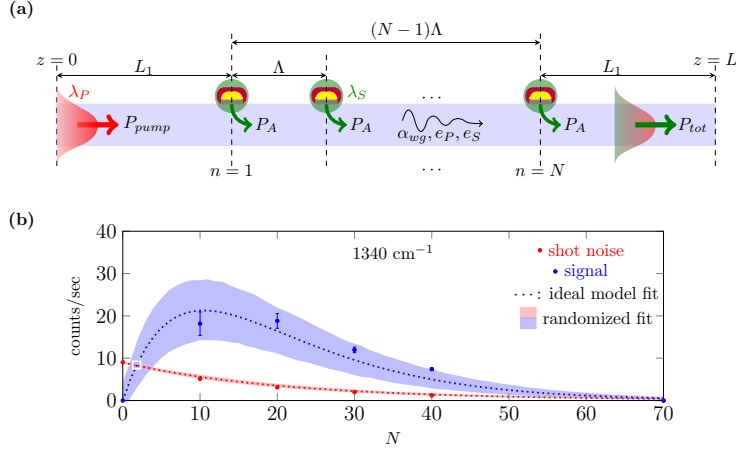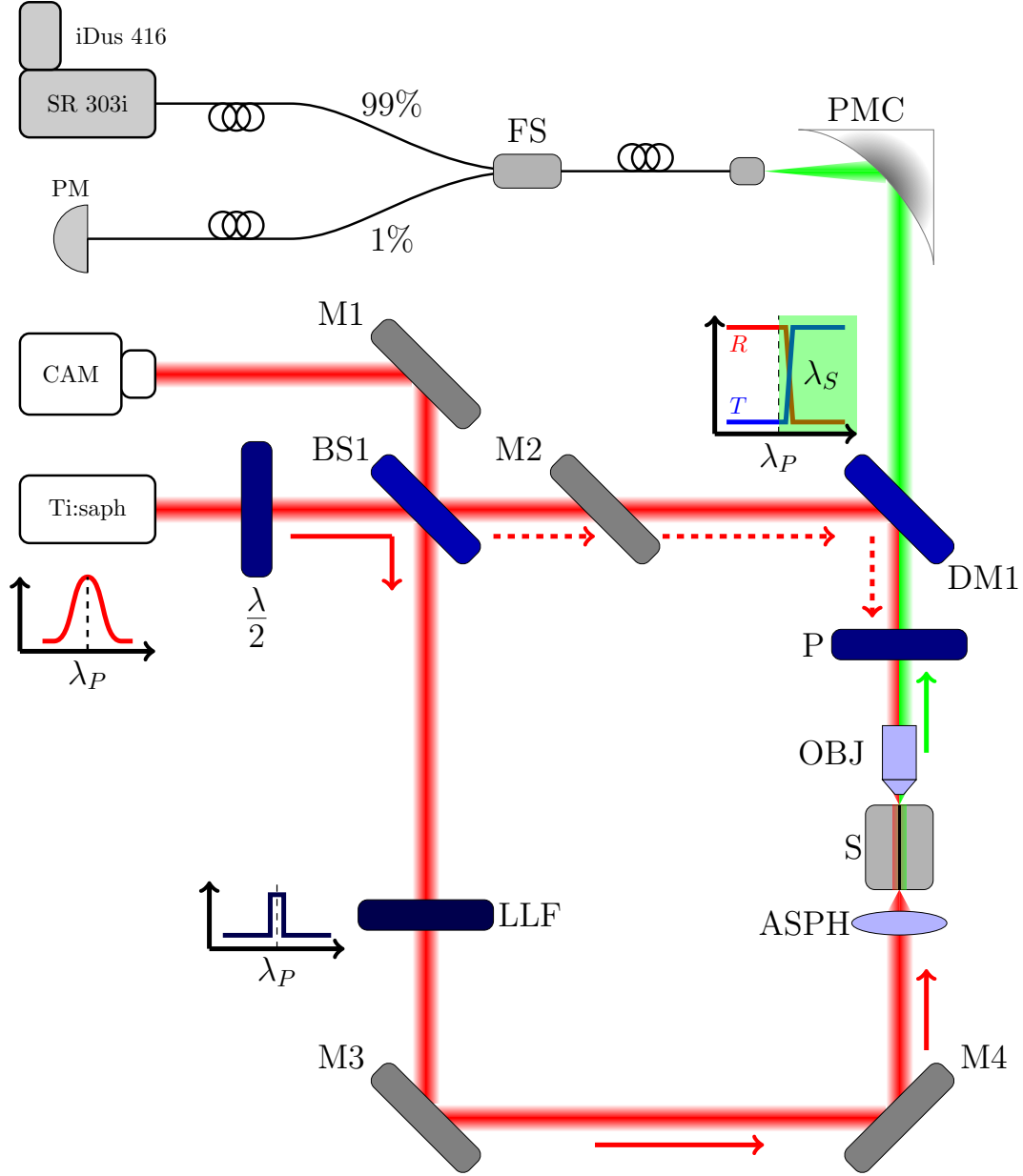
15

(a)

(b)

Figure 4: **On-chip SERS model and fit to the experimental data. (a)** Longitudinal cross-section of the chip. The pump beam, with power P_{pump} at wavelength λ_P, excites a Stokes signal λ_S of which the total power at the output facet is P_{tot}. The plasmonic array consists of N antennas with period Λ and is separated L_1 from both the input and output facet of the chip. Each of the antennas generates P_A guided Stokes power for a given pump power. Apart from the waveguide losses α_{wg}, the pump and Stokes light is attenuated by the pump e_P and Stokes e_S extinction respectively. **(b)** Signal (blue circles) and shot noise (red circles) at the 1,340 cm^{-1} peak. The dotted lines represent a constrained fit to the ideal model while the shaded areas represent a randomized fit to estimate the uncertainty on the experimental parameters. The white square denotes the minimum number of antennas N_{min} required to generate a detectable signal.

Figure 4

Supporting Information
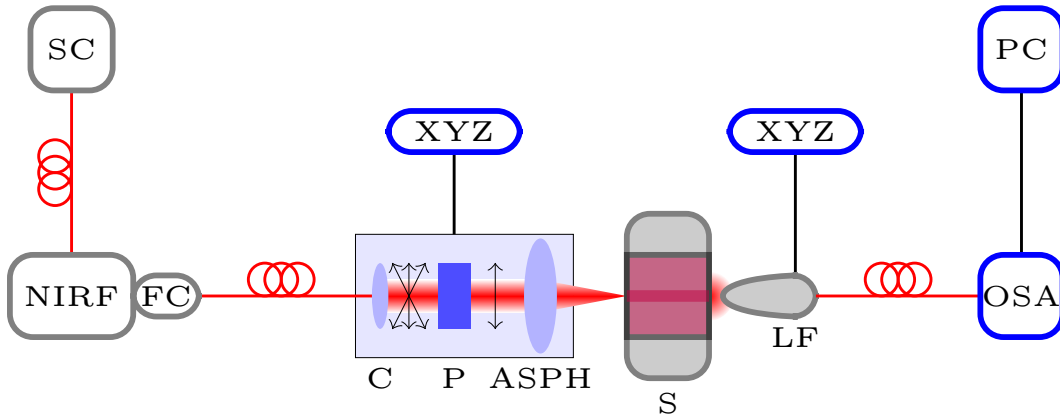
S1: SERS measurement setup

SERS spectra are measured with the setup depicted in Figure S1. A tunable Ti:saph laser is set to a pump wavelength of $\lambda_P = 785$ nm (red) after which the polarized beam passes through a half-wave plate ($\lambda/2$) in order to rotate the polarization to a TE-polarized beam. A beamsplitter BS1 then splits the beam into two parts (solid and dashed red line). The solid path is used to generate the forward propagating Raman beam and passes through a laser line filter (LLF) at λ_P for side-band suppression before it is coupled into the chip by an aspheric lens (ASPH). The output beam is then collected with an objective (OBJ) and passes through a polarizer P (set to TE) before it is filtered by a dichroic mirror which reflects the pump beam and transmits all Stokes wavelengths (green). The Stokes light is collected into a fiber using a parabolic mirror collimator (PMC) after which the fiber is split by a fiber splitter (FS) of which 1% goes to a power meter (PM) and 99% to a commercial spectrometer from ANDOR (Shamrock 303i spectrometer and iDus 416 cooled CCD camera). Mirror M2 blocks the second (dashed) path during the measurement but can be removed for alignment purposes. The camera (CAM) and the 1% fiber tap are used during alignment and to measure the transmitted power P_T. In order to align the sample we initially set the wavelength of the Ti:saph to $\lambda_T = 800$ nm (such that it can be transmitted through DM1 and collected in the power meter) and maximize the transmitted power P_T in the forward path. After optimizing the transmission the laser is tuned back to $\lambda_P = 785$ nm.

Supplementary Figure S1: **Measurement setup.** Ti:saph: tunable Ti:saphire laser emitting the pump beam at $\lambda_P = 785$ nm, CAM: camera, PM: power meter, SR 303i and iDus 416: spectrometer and cooled CCD detector from ANDOR, BS1: beamsplitter, $\lambda/2$: half-wave plate, LLF: laser line filter for 785 nm, P: polarizer, M1/M3/M4: fixed mirrors, M2: removable mirror/beam block, OBJ: objective (50X, NA=0.9), ASPH: aspheric lens (NA=0.5), S: sample stage, DM1: dichroic mirror (reflection R and transmission T shown), PMC: parabolic mirror collimator (EFL=15 mm, NA=0.2), FS: fiber splitter.
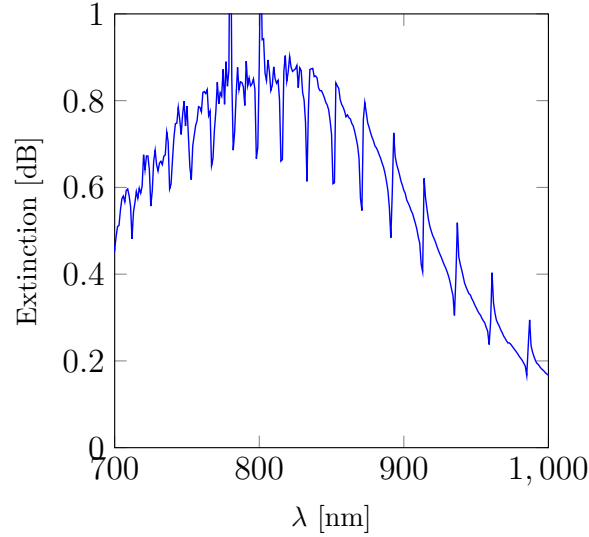
S2: Extinction measurement

Single antenna extinction spectra, resulting from the plasmon resonance, are measured with the setup depicted in Figure S2. Light from an NKT EXR-4 supercontinuum source (SC) is filtered through a near-IR filter (NIRF) to filter out the relevant wavelength region. Subsequently it is coupled in a fiber with a fiber coupling unit (FC). This fiber is plugged into a fiberbench consisting of 3 parts (fixed to the same bench): an achromatic fiber collimator (C) which converts the fiberized light to a free-space beam, a free-space broadband polarizer (P) which polarizes the unpolarized light into a TE-beam and an aspheric lens (ASPH) used to focus the free-space beam on the input facet of the chip. This fiberbench (marked by the light-blue area) is mounted on a piezo-controlled stage (XYZ) in order to precisely couple the supercontinuum light into the chip. At the output facet a lensed fiber (LF) is used to capture the transmitted light. This lensed fiber is also connected to a piezo-controller for accurate positioning. Finally the light is coupled to an optical spectrum analyzer (OSA) and the spectra are read out by Python controlled software (PC).

Supplementary Figure S2: **Resonance measurement.** SC: supercontinuum source, NIRF: near-IR filter, FC: fiber coupling unit, C: achromatic fiber collimator, P: free-space broadband polarizer, ASPH: aspheric lens (NA=0.68), LF: lensed fiber, XYZ: piezo controller stage, S: sample stage, OSA: Optical Spectrum Analyzer, PC: OSA control using Python based measurement framework.

The single antenna extinction curves $E(\lambda)$ (in dB) can then be calculated through $E(\lambda) = (T_{ref} - T_{Nant})/N$ in which T_{ref} is the power transmission (in dB) through the reference waveguide and T_{Nant} the power transmission (in dB) of a waveguide functionalized with N antennas.[16] In Figure S3 the single antenna extinction curve of a waveguide functionalized with $N = 40$ bowtie antennas ($\alpha \approx 60°$, $H \approx 100$ nm and $\Delta \approx 40$ nm) is shown. The extinction exhibits periodic fringes on the broad envelope which are attributed to interference effects of the plasmonic array. On one hand, the array forms a multiple Fabry-Perot resonator of which the expected free spectral range FSR of 19.9 nm around 1,250 cm^{-1} (using the fabricated array period of $\Lambda = 10\mu$m) matches well with the experimentally obtained value (≈ 20.8 nm). On the other hand, the waveguide mode interferes with the radiative decay of the plasmon mode. This will affect the specific lineshape and strength of the fringes. The spectral positions at which we see sudden changes in the Raman background (see main text) coincide with the fringes observed on the extinction curves. Hence these features are not attributed to specific Raman lines.

Supplementary Figure S3: **Single antenna extinction spectrum.** Single antenna extinction spectrum of the waveguide functionalized with $N = 40$ bowtie antennas ($\alpha \approx 60°$, $H \approx 100$ nm and $\Delta \approx 40$ nm).

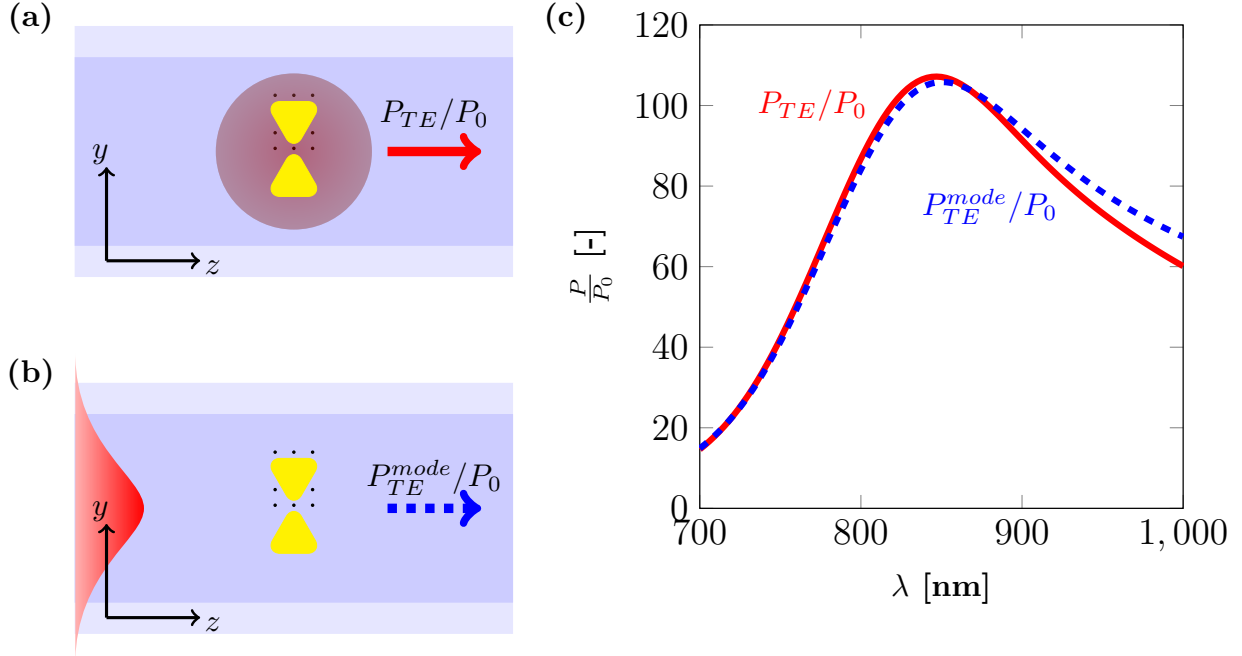## S3: Derivation of the analytical on-chip SERS model

The power $P_{wg}(\mathbf{r}_0, \lambda)$ coupled into the forward propagating waveguide mode $\mathbf{E}^m(\mathbf{r}, \lambda)$ as a result of a radiating dipole at position $\mathbf{r}_0$ is given by

$$\frac{P_{wg}(\mathbf{r}_0, \lambda)}{P_0} = \frac{3}{8\pi} \frac{n_g(\lambda)}{n_m} \left(\frac{\lambda}{n}\right)^2 \frac{\varepsilon_o \varepsilon(\mathbf{r}_0) \left|\mathbf{e}_d \cdot \mathbf{E}^m(\mathbf{r}_0, \lambda)\right|^2}{\int \int_{A_{wg}} \varepsilon_o \varepsilon(\mathbf{r}) \left|\mathbf{E}^m(\mathbf{r}, \lambda)\right|^2 d\mathbf{r}} \bigg|_{z=z_0} \tag{1}$$

where $P_0 = \frac{\omega^4 |\mathbf{d}_0|^2}{12\pi\varepsilon_0 c^3}$ is the power radiated in free space at a wavelength $\lambda = \frac{2\pi c}{\omega}$, $\varepsilon(\mathbf{r})$ the relative permittivity at position $\mathbf{r}$, $n_g(\lambda)$ the group index of the waveguide mode, n_m the refractive index of the medium in which the dipole is placed and $\mathbf{e}_d$ the unit vector along the dipole moment vector $\mathbf{d}_0$.[31] The integral in the denominator is calculated over a waveguide cross-section A_{wg} in the xy-plane and evaluated at the dipole position $z = z_0$ (coordinate axes are defined in Figure S4(a) and S4(b)).

Functionalizing the waveguide with an antenna however introduces a perturbation to the modal field $\mathbf{E}^m(\mathbf{r}, \lambda)$ of the waveguide. In order to investigate whether formula (1) is still valid when the waveguide is functionalized with a metallic antenna, two sets of simulations were performed . In the first set (Figure S4(a)) the coupling of dipole radiation into the fundamental TE-mode is investigated. This allows an explicit evaluation of the left-hand side of equation (1). Dipole sources with fixed dipole moment vector $\mathbf{d}_0$ were placed at different positions $\mathbf{r}_0$ around the antenna (shown as black dots in Figure S4(a)). In the second set (Figure S4(b)) the fundamental TE-mode is launched into the waveguide functionalized with the same antenna in order to calculate the fields at positions $\mathbf{r}_0$ where the dipole sources in the first simulation set were located. In this way it is possible to calculate the right-hand side of equation (1). The integral in the denominator is approximated in all calculations by evaluating it on a reference waveguide without antenna, so using the true modal fields and not the perturbed fields. This approximation allows a sufficient accurate evaluation of the right-hand side of equation (1) as is confirmed in Figure S4(c) where the simulation results are depicted. The red curve represents an incoherent superposition of the power coupled into the TE-mode as a result of the complete set of dipole emitters. The dashed

blue curve is the predicted power that couples into the TE-mode, and is calculated as an incoherent superposition of the predicted coupled powers for each dipole. It is clear that the predicted power coupled into the TE-mode (simulation set 2) matches very well with the explicit calculation using dipole emitters (simulation set 1).

Supplementary Figure S4: **Dipole radiation.** **(a)** Simulation set 1: collection of incoherent dipoles emitting into the fundamental TE-mode (red). **(b)** Simulation set 2: Calculation of the electric fields at the dipole positions of simulation set 1 using a fundamental TE-mode excitation. Based on these fields one can extract the predicted dipole emission (dashed blue). **(c)** Simulation results: incoherent superposition of the TE-coupled power due to all dipoles (red) and predicted TE-coupled power using the electric field values (dashed blue).

Since the right-hand side of equation (1) faithfully represents the power coupled into the fundamental TE-mode, it is used to derive an analytical model predicting the total Stokes power coupled into the waveguide as a result of an array of N antennas on top of the waveguide (see Figure 4(a) in the main text). First of all we write the dipole strength $|\mathbf{d}_0|^2$ as a function of the guided pump power P_{pump} in the waveguide

$$|\mathbf{d}_0|^2 = \alpha_m^2 \, |\mathbf{E}_d^m(\mathbf{r}_0, \lambda)|^2 \, \frac{n_g(\lambda) P_{pump}}{\int\int c\varepsilon_o\varepsilon(\mathbf{r}) \, |\mathbf{E}^m(\mathbf{r}, \lambda)|^2 \, d\mathbf{r} \, |_{z=z_{input}}}$$

in which α_m is the molecular polarizability and $\mathbf{E}_d^m(\mathbf{r}_0, \lambda) = \mathbf{e}_d \cdot \mathbf{E}^m(\mathbf{r}_0, \lambda)$ the field strength at the dipole position. The integral is evaluated at the input $z_{input} = 0$ of the waveguide.[24] The quantity

$$\eta(\mathbf{r}_0, \lambda_P, \lambda_S) = \frac{n_g(\lambda_P)n_g(\lambda_S)\lambda_S^2}{n_m} \frac{|\mathbf{E}_d^m(\mathbf{r}_0, \lambda_P)|^2 |\mathbf{E}_d^m(\mathbf{r}_0, \lambda_S)|^2}{\left(\int\int \varepsilon(\mathbf{r}) |\mathbf{E}^m(\mathbf{r}, \lambda_P)|^2 d\mathbf{r}\right)\left(\int\int \varepsilon(\mathbf{r}) |\mathbf{E}^m(\mathbf{r}, \lambda_S)|^2 d\mathbf{r}\right)}. \quad (2)$$

solely depends on the modal fields and can be evaluated on a reference waveguide. Both integrals in the denominator are evaluated along the same cross-sectional area of the reference waveguide (in all simulations $n_m = 1$). The power P_{wg}^n coupled into the fundamental TE-mode due to a collection of incoherently radiating dipoles around antenna n ($n = 1 \ldots N$) is then approximately given by

$$\frac{P_{wg}^n}{P_{pump}} \approx \frac{\rho \sigma(\lambda_S)}{2} e^{-\alpha(\lambda_P)(L_1 + (n-1)\Lambda)} e(\lambda_P)^{1-n} \int\int\int_{V_m} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) \beta(\mathbf{r}_0, \lambda_P)^2 \beta(\mathbf{r}_0, \lambda_S)^2 d\mathbf{r}_0 \quad (3)$$

in which ρ is the molecular density, $\sigma(\lambda_S) = \pi^2 \alpha_m^2/(\varepsilon_0^2 \lambda_S^4)$ the Raman cross section,[32] $\alpha(\lambda_P) = \alpha_P$ the waveguide loss at the pump wavelength, $e(\lambda_P) = e_P$ the linear antenna extinction at the pump wavelength, V_m the volume in which the molecules are situated and $\beta(\mathbf{r}_0, \lambda_{P/S}) = \beta_{P/S}(\mathbf{r}_0)$ the field enhancement factor, at the dipole position $\mathbf{r}_0$, at the pump (P) and Stokes (S) wavelength respectively:

$$\beta(\mathbf{r}_0, \lambda) = \frac{|\mathbf{E}_d^{ant}(\mathbf{r}_0, \lambda)|}{|\mathbf{E}_d^m(\mathbf{r}_0, \lambda)|}$$

whereby $\mathbf{E}_d^{ant}(\mathbf{r}_0, \lambda)$ is the local field around the antenna surface. For notational simplicity we rewrite formula (3) as

$$\frac{P_{wg}^n}{P_{pump}} \approx P_A(\lambda_P, \lambda_S) e^{-\alpha_P(L_1 + (n-1)\Lambda)} e_P^{1-n}$$

in which

$$P_A(\lambda_P, \lambda_S) = \frac{\rho \sigma(\lambda_S)}{2} \int\int\int_{V_m} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) \beta(\mathbf{r}_0, \lambda_P)^2 \beta(\mathbf{r}_0, \lambda_S)^2 d\mathbf{r}_0$$

is a dimensionless antenna dependent factor which incorporates the specific field enhancement profile near the antenna surface. Formula (3) takes into account the waveguide loss and the loss induced by the antennas in front of the nth antenna as a result of which the actual excitation power

decays. The signal (at Stokes wavelength λ_S) still has to pass $N-n$ antennas and propagate along a distance $L_1 + (N-n)\Lambda$ such that the power $P_{wg,out}^n$ reaching the output of the waveguide is given by

$$\frac{P_{wg,out}^n}{P_{pump}} \approx P_A(\lambda_P, \lambda_S) \left(e^{-\alpha_P(L_1+(n-1)\Lambda)} e_P^{1-n} \right) \left(e^{-\alpha_S(L_1+(N-n)\Lambda)} e_S^{n-N} \right).$$

$$= P_A(\lambda_P, \lambda_S) \left(e^{-\alpha_P(L_1-\Lambda)-\alpha_S(L_1+N\Lambda)} e_P e_S^{-N} \right) \left(\left(\frac{e_S}{e_P} \right) e^{(\alpha_S-\alpha_P)\Lambda} \right)^n.$$

Since N antennas contribute incoherently to the signal, the total amount of Stokes light at the output of the waveguide is given by

$$\frac{P_{wg,out}^{tot}}{P_{pump}} \approx P_A(\lambda_P, \lambda_S) \left(e^{-\alpha_P(L_1-\Lambda)-\alpha_S(L_1+N\Lambda)} e_P e_S^{-N} \right) \sum_{n=1}^{N} \left(\left(\frac{e_S}{e_P} \right) e^{(\alpha_S-\alpha_P)\Lambda} \right)^n.$$

For the considered waveguide platform it is reasonable to approximate $\alpha_P \approx \alpha_S = \alpha_{wg}$ such that the forward propagating Raman power is given by

$$\frac{P_{wg,out}^{tot}}{P_{pump}} = \frac{P_{tot}}{P_{pump}}$$

$$\approx P_A(\lambda_P, \lambda_S) \left(e^{-\alpha_{wg}(2L_1+(N-1)\Lambda)} e_P e_S^{-N} \right) \sum_{n=1}^{N} \left(\frac{e_S}{e_P} \right)^n$$

$$\approx P_A(\lambda_P, \lambda_S) \left(e^{-2\alpha_{wg}L_1} e_S^{1-N} \right) \left(\frac{1 - \left(\frac{e_S}{e_P} \right)^N}{1 - \left(\frac{e_S}{e_P} \right)} \right) = FOM(N, \lambda_P, \lambda_S) e^{-2\alpha_{wg}L_1},$$

where $L = 2L_1 + (N-1)\Lambda \approx 2L_1$ has been used. This result applies to Stokes light co-propagating with the pump beam. For a fixed antenna geometry and λ_P and λ_S, the optimum number of antennas N_{opt} that should be patterned on a waveguide to maximize the SERS signal is given by

$$N_{opt} = \frac{\log \left(\frac{\log(e_S)}{\log(e_P)} \right)}{\log \left(\frac{e_S}{e_P} \right)}.$$

For the specific case of a monolayer, one has a molecular surface density ρ_s (number of molecules

per m^2) rather than a volume density ρ (number of molecules per m^3). In this limiting case the single antenna conversion efficiency is theoretically defined by

$$P_A(\lambda_P, \lambda_S) = \frac{\rho_s \sigma(\lambda_S)}{2} \int \int_{A_m} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) \beta(\mathbf{r}_0, \lambda_P)^2 \beta(\mathbf{r}_0, \lambda_S)^2 d\mathbf{r}_0$$

where the integration is now performed over the surface area A_m covered by the monolayer. Numerically one however always needs to model the presence of such a monolayer by introducing a finite thickness t_m. As a result an effective monolayer volume $V_m \approx t_m \times A_m$ is defined. For any numerical evaluation the fields are hence integrated over this volume V_m such that

$$\int \int \int_{V_m} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) \beta(\mathbf{r}_0, \lambda_P)^2 \beta(\mathbf{r}_0, \lambda_S)^2 d\mathbf{r}_0 \approx t_m \times \int \int_{A_m} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) \beta(\mathbf{r}_0, \lambda_P)^2 \beta(\mathbf{r}_0, \lambda_S)^2 d\mathbf{r}_0$$

The single antenna conversion efficiency is then given by

$$P_A(\lambda_P, \lambda_S) \approx \frac{\rho_s \sigma(\lambda_S)}{2t_m} \int \int \int_{V_m} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) \beta(\mathbf{r}_0, \lambda_P)^2 \beta(\mathbf{r}_0, \lambda_S)^2 d\mathbf{r}_0$$

From a numerical point of view, an effective volume density $\rho_{eff} = \rho_s/t_m$ is introduced which is then multiplied with the integral over the effective monolayer volume V_m.

In order to correlate the Stokes power at the output facet of the chip and the actual detected signal counts C_S, normalized with the integration time T, we need to take into account the characteristics of all optics in our setup. The Stokes power reaching the detector surface is given by

$$P_S = \gamma^{out} \times T_{out} \times P_{tot} = \gamma^{out} \times T_{out} \times \left(FOM e^{-\alpha_{wg}L} P_{pump}\right)$$
$$= \gamma^{out} \times T_{out} \times \left(FOM e^{-\alpha_{wg}L} \left(\gamma^{in} P_{in}\right)\right)$$
$$= FOM \left(\gamma^{in} \gamma^{out} P_{in} e^{-\alpha_{wg}L}\right) \times T_{out}$$

in which γ^{in} and γ^{out} are the coupling efficiencies in and out of the chip respectively, P_{in} is the power just before the input facet of the chip and T_{out} is the transmission through all optics between the output of the chip and the detector surface. The quantity $\gamma^{in}\gamma^{out}P_{in}e^{-\alpha_{wg}L}$ can be written as

$$\gamma^{in}\gamma^{out}P_{in}e^{-\alpha_{wg}L} = P_T/T_{PM}$$

in which P_T is the transmitted power as measured by the power meter (see Supplementary Information S1) and T_{PM} is the optical transmission between the output of the chip and the input of the powermeter. This transmission is related to T_{out} by $T_{out} \approx 100 \times T_{PM} \times T_{spec}$ in which T_{spec} is the transmission between the input slit of the spectrometer and the detector surface. The factor 100 stems from the fact that the fiber splitter only transmits 1% of the power to the power meter. So eventually P_S can be written as $P_S = FOM \times P_T \times 100 T_{spec}$. By using the spectrometer sensitivity χ (defined as the number of electrons per count) and quantum efficiency $QE(\lambda_S)$ (number of electrons per number of incident photons) P_S can be related to the measured signal counts C_S per unit integration time T by

$$P_S = \left(\frac{C_S}{T}\right)\left(\frac{hc}{\lambda_S}\right)\left(\frac{\chi}{QE(\lambda_S)}\right).$$

Finally one gets

$$\frac{C_S}{T} = FOM \times 100 T_{spec} \left(\frac{\lambda_S}{hc}\right)\left(\frac{QE(\lambda_S)}{\chi}\right) P_T = FOM \times C^*(\lambda_S).$$

The transmitted power $100 P_T$ is about 1mW for the given measurement setup. If we consider the 1340 cm^{-1} line then $\frac{QE(\lambda_S)}{\chi} \approx 1$ counts/photon and $T_{spec} \approx 0.44$ (using data supplied by the manufacturer) such that $C^* \approx 1.94 \times 10^{15}$ counts/sec. $C^*(\lambda_S)$ can now be used as a conversion factor when the experimental signal counts are fitted to the analytical model (and hence the figure of merit). In this way one can rigorously quantify the parameters e_P, e_S and P_A and compare them with the theoretical predictions.

S4: Derivation of the background signal

Similar reasonings as in the Supplementary Information S3 can be applied to derive the total background signal generated along the waveguide. This background mostly stems from the SiN core and is hence generated when the pump light propagates along the total length L of the waveguide. For a functionalized waveguide the attenuation of the pump and Stokes light due to the antenna array also needs to be incorporated. The dipoles which give rise to the background are mainly situated in the core, so for a given core cross-sectional area A_{core} equation (2) has to be integrated over the complete waveguide core. This quantity is defined as

$$\eta_{core}(\lambda_P, \lambda_S) = \eta_c = \int \int_{A_{core}} \eta(\mathbf{r}_0, \lambda_P, \lambda_S) d\mathbf{r}_0.$$

For a given molecular core density ρ_c and core scattering cross section σ_c, the total background P_{bg} propagating in the forward direction is then calculated through

$$\frac{P_{bg}}{P_{pump}} = \frac{\rho_c \sigma_c \eta_c}{2} \left\{ \left(\int_0^{L_1} e^{-\alpha_P z} e^{-\alpha_S(L-z)} dz \right) e_S^{-N} + \left(\int_{L_1+(N-1)\Lambda}^{2L_1+(N-1)\Lambda} e^{-\alpha_P z} e^{-\alpha_S(L-z)} dz \right) e_P^{-N} + \right.$$
$$\left. \sum_{n=0}^{N-2} \left(\int_{L_1+n\Lambda}^{L_1+(n+1)\Lambda} e^{-\alpha_P z} e^{-\alpha_S(L-z)} dz \right) e_P^{-(n+1)} e_S^{-(N-n-1)} \right\}.$$

After some calculation and using $\alpha_P \approx \alpha_S = \alpha_{wg}$ and $L \approx L_1/2$ ($2L_1 >> (N-1)\Lambda$), the total forward propagating background is given by

$$\frac{P_{bg}}{P_{pump}} = \frac{\rho_c \sigma_c \eta_c}{2} e^{-\alpha_{wg}L} \left\{ (e_P^{-N} + e_S^{-N}) \frac{L}{2} + \Lambda \frac{e_S^{1-N}}{e_P} \left(\frac{1 - \left(\frac{e_S}{e_P} \right)^{N-1}}{1 - \left(\frac{e_S}{e_P} \right)} \right) \right\}$$
$$\approx \frac{\rho_c \sigma_c \eta_c}{2} e^{-\alpha_{wg}L} \left\{ (e_P^{-N} + e_S^{-N}) \frac{L}{2} \right\}$$
$$= P_B(\lambda_P, \lambda_S) e^{-\alpha_{wg}L} \left(e_P^{-N} + e_S^{-N} \right) \approx P_B(\lambda_P, \lambda_S) e^{-2\alpha_{wg}L_1} \left(e_P^{-N} + e_S^{-N} \right).$$

S5: Randomized fit to a generalized model

The model derived in the Supplementary Information S3 and S4 assumes identical antennas. Due to fabrication errors there will always be differences among each of the antennas in the array. These differences have an impact on both the extinction e_P and e_S as well as on the single antenna conversion efficiency $P_A(\lambda_P, \lambda_S)$. While the experimental data described in this Letter can be fitted well to the ideal model, the fit is not perfect. In this section a randomized fit model that takes into account the potential deviations among different antennas is outlined. The power generated by antenna n is given by:

$$\frac{P_{wg}^n}{P_{pump}} \approx P_A(\lambda_P, \lambda_S) e^{-\alpha_P(L_1+(n-1)\Lambda)} e_P^{1-n}.$$

Since $L_1 >> N\Lambda$ this can be simplified to

$$\frac{P_{wg}^n}{P_{pump}} \approx P_A(\lambda_P, \lambda_S) e^{-\alpha_P L_1} e_P^{1-n}.$$

The formula however assumes a constant pump extinction from the previous $n-1$ antennas. Furthermore the factor $P_A(\lambda_P, \lambda_S)$ is also assumed constant for each antenna. If we allow that each of the antennas has a different extinction e_P^m and antenna dependent factor $P_A^m(\lambda_P, \lambda_S)$ ($m = 1 \ldots N$), then the power generated by antenna n is given by

$$\frac{P_{wg}^n}{P_{pump} e^{-\alpha_P L_1}} \approx P_A^n(\lambda_P, \lambda_S) \prod_{m=1}^{n-1} (e_P^m)^{-1}.$$

Applying a similar reasoning to the Stokes light that has to propagate along $N-n$ other antennas (and a length L_1), the Stokes power (generated by antenna n) reaching the output is given by

$$\frac{P_{wg,out}^n}{P_{pump} e^{-\alpha_P 2L_1}} \approx P_A^n(\lambda_P, \lambda_S) \left(\prod_{m=1}^{n-1} (e_P^m)^{-1} \right) \left(\prod_{m=n+1}^{N} (e_S^m)^{-1} \right).$$

The total generated Stokes power from all N (potentially different) antennas is then finally

$$\frac{P_{tot}}{P_{pump}e^{-2\alpha_{wg}L_1}} \approx \sum_{n=1}^{N} P_A^n(\lambda_P, \lambda_S)\left(\prod_{m=1}^{n-1}(e_P^m)^{-1}\right)\left(\prod_{m=n+1}^{N}(e_S^m)^{-1}\right). \tag{4}$$

For the background one can analogously write

$$\frac{P_{bg}}{P_{pump}e^{-2\alpha_{wg}L_1}} \approx P_B\left(\prod_{m=1}^{N}(e_P^m)^{-1} + \prod_{m=1}^{N}(e_S^m)^{-1}\right) \tag{5}$$

In this case P_B is a constant since it only depends on the waveguide parameters. The shot noise associated to the background is then simply proportional to $\sqrt{P_{bg}}$. The number of counts/sec can be obtained by multiplying equations (4) and (5) with $C^*(\lambda_S)$ (see Supplementary Information S3). Based on the conversion factor $C^*(\lambda_S)$ and a fit to the experimental data, one can quantitatively determine a value for the parameters in the model. For a fixed λ_P and λ_S, $P_A(\lambda_P, \lambda_S)C^*(\lambda_S)$ will be denoted by P_A^*.

In order to make a randomized fit of the signal and shot noise at the same time, N random numbers (based on a normal distribution) are independently generated for the pump and Stokes extinction and for the single antenna conversion efficiency (so three random numbers are generated for each antenna in the array). Using these random numbers, formulas (4) and (5) are evaluated. This process is repeated 1000 times in order to generate a statistically relevant distribution of the possible signal and shot noise counts. Ultimately the mean value μ and standard deviation σ from the obtained signal and shot noise distributions are extracted for each number of antennas N. The 3σ-intervals for both distributions ($[\mu - 3\sigma, \mu + 3\sigma]$) define an area that marks the possible signal and noise counts for a given uncertainty on the antenna parameters (according to formulas (4) and (5)). These areas are plotted in Figure 4(b) of the main text.

The mean value of each of the three normal distributions (for e_P, e_S and P_A) is determined by a fit of the experimental data to the ideal model. One could obtain a perfect fit to either the signal or shot noise data by fitting only one of the two equations ((4) or (5)). The fitting parameters would however not generate a perfect fit to the other non-fitted equation since both equations depend on

e_P and e_S. Therefore a constrained fit, in which one minimizes the fitting error to both equations simultaneously, is performed such that both the signal and shot noise data are represented well using the fitted values for e_P and e_S. The values e_P, e_S and P_A^* (mind that P_A^* and not P_A is used because we fit the counts/sec) obtained from the constrained fit are then chosen to be the mean values (μ_{e_P}, μ_{e_S} and $\mu_{P_A^*}$) of their respective normal distributions. The standard deviation σ_x is chosen such that the $3\sigma_x$ intervals represent realistic deviations from the mean value (x denotes one the three parameters). For the extinction this means e.g. that $\mu_{e_{P/S}} - 3\sigma_{e_{P/S}} \geq 1$ (since 1 is the lower boundary for the linear extinction). Since P_A^* is always positive, $\mu_{P_A^*} - 3\sigma_{P_A^*} > 0$ should also be satisfied. For the experimental data shown in Figure 3 of the main text, $\mu_{e_P} = 1.1182$, $\mu_{e_S} = 1.0845$ and $\mu_{P_A^*} = 5.0394$ has been derived from the constrained fit. The sigma value for the extinction is chosen to be $\sigma = 0.028$ such that $E_P \approx 0.49 \pm 0.11$ dB and $E_S \approx 0.35 \pm 0.11$ dB. For P_A^* we similarly get $P_A^* \approx 5.04 \pm 1.5$ such that the single antenna conversion efficiency for the 1340 cm^{-1} line is $P_A = (2.60 \pm 0.77) \times 10^{-15}$. All of these values satisfy the constraints mentioned above and represent realistic deviations of the three fitting parameters.